Epipolar rectification

Pascal Monasse, Jean-Michel Morel, Zhongwei Tang

October 29, 2010

Overview.

Rectification example: original images

Rectification example: original images

Rectification example: rectified images

Rectification example: rectified images

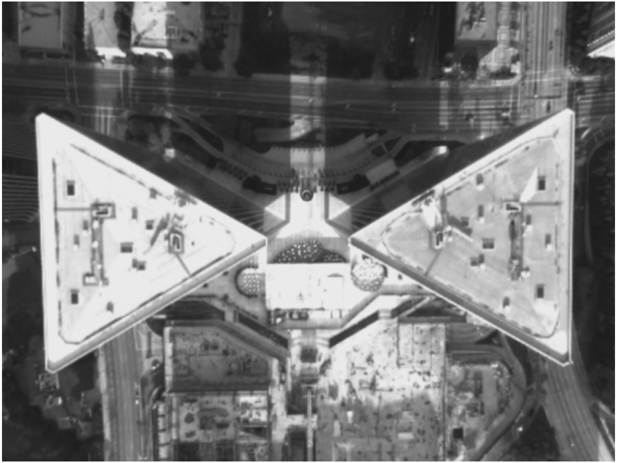
Rectification example: original images

Rectification example: original images

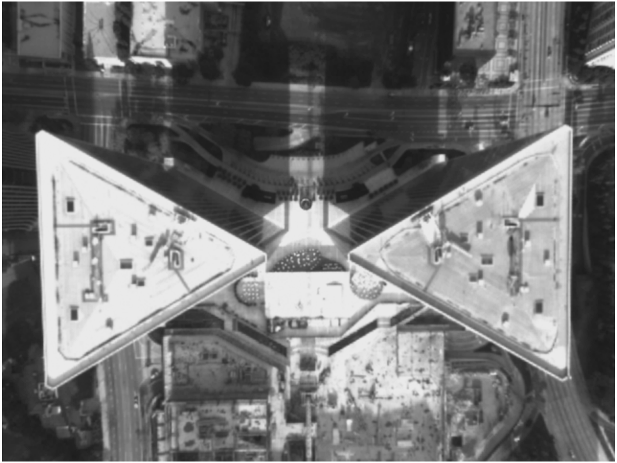
Rectification example: rectified images

Rectification example: rectified images

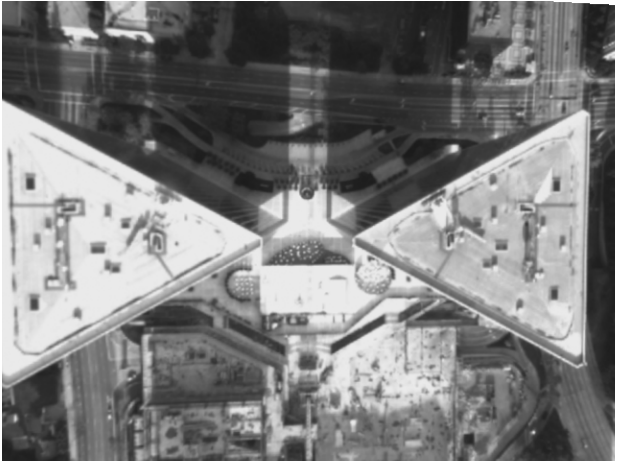
Rectification example: original images



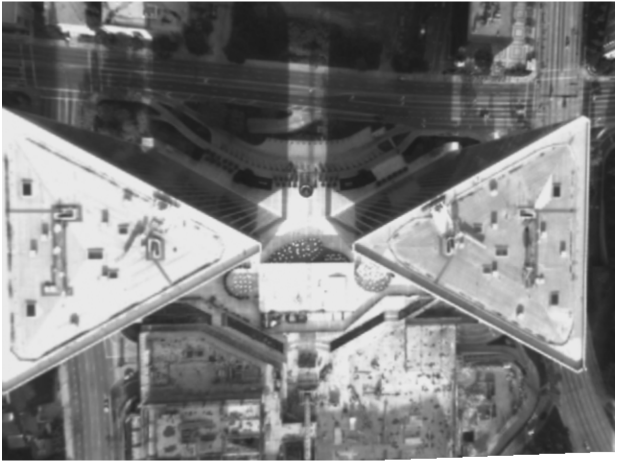
Rectification example: original images



Rectification example: rectified images (failed)

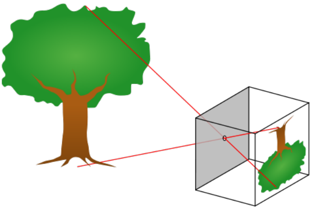


Rectification example: rectified images (failed)



Pinhole camera model

- An ideal model where the camera aperture is described as a point and no lenses are used to focus light
- No geometric distortions or blurring of unfocused objects caused by lenses and finite sized apertures
- The mathematical relationship between the coordinates of a 3D point and its projection onto the image plane



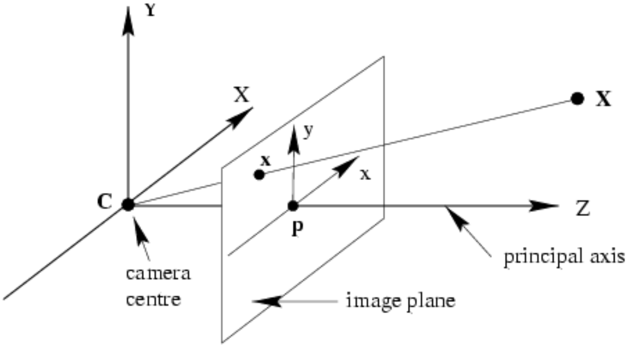


Figure: Pinhole camera model.

Terminology

- camera center (optic center): the point by which all the rays pass.
- image plane: the camera CCD plane where the image is formed.
- principal axis: the line from the camera center perpendicular to the image plane.
- principal plane: the plane containing the camera center and parallel to the image plane.
- world frame: a pre-fixed frame where any 3D point can be represented.
- camera frame: the frame based on camera which has camera center as origin and principal axis as Z-axis.
- focal length f: the distance from the camera center to the image plane.

Central projection

Definition

A mapping of 3D space into a plane P that associates with any point the intersection with P of the line passing through the point and a fixed point.

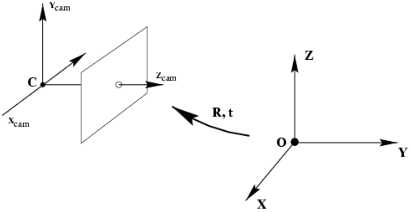
The pinhole camera is a central projection.

Frame change

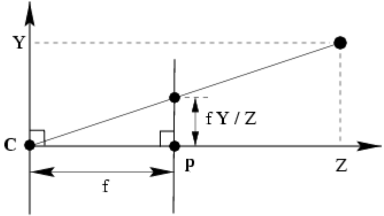
Represent a 3D point in the camera frame by a translation and a rotation from the world frame:

$$\hat{\mathbf{X}}_c = \mathbf{R}(\hat{\mathbf{X}} - \mathbf{C}) \quad (1)$$

with $\hat{\mathbf{X}} = (X, Y, Z)^T$ and $\hat{\mathbf{X}}_c = (X_c, Y_c, Z_c)^T$ the coordinate of a point in the world frame and in the camera frame respectively; $\mathbf{C} = (X_o, Y_o, Z_o)^T$ the camera center in the world frame; $\mathbf{R}$ the rotation from world frame to the camera frame.



Central projection



$\hat{\mathbf{X}}_c$ is projected to the point $\mathbf{x}_c = (x_c, y_c)^T$ on the image plane (Thales's theorem):

$$x_c = fX_c/Z_c \quad (2)$$
$$y_c = fY_c/Z_c \quad (3)$$

Matrix form

More succinct in matrix form by using homogeneous coordinate:

$$\mathbf{x}_c = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \hat{\mathbf{X}}_c \tag{4}$$

In homogeneous coordinate, $\mathbf{x}_c = (fX_c, fY_c, Z_c)^T$ is equivalent to the 2D point $(fX_c/Z_c, fY_c/Z_c)^T$ by dividing the first two coordinates by the third coordinate.

All in matrix form

By concatenating the frame change and the central projection, a 3D point is projected to a 2D point:

$$\mathbf{x}_c = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \hat{\mathbf{X}}_c = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{R} \left[\mathbf{I} \mid -\mathbf{C} \right] \begin{pmatrix} \hat{\mathbf{X}} \\ 1 \end{pmatrix} \tag{5}$$

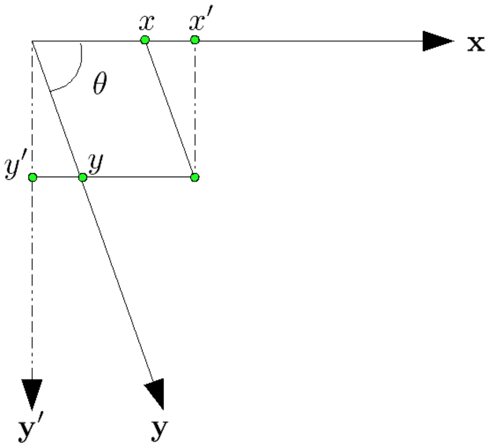
$$= \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{R} \left[\mathbf{I} \mid -\mathbf{C} \right] \mathbf{X}$$

with $\mathbf{X} = (X, Y, Z, 1)^T$ the homogeneous coordinates of 3D point $\hat{\mathbf{X}} = (X, Y, Z)^T$.

CCD plane in pixels

- The above obtained 2D point $\mathbf{x}_c$ has the unity in meter or millimeter. But any digital image is measured in the unity of pixels.
- $\mathbf{x}_c$ has the principal point as the origin, while the convention is to take top-left corner of image as the origin.
- Due to some manufacture imprecision, the CCD array is not exactly a rectangular grid.

skewness



$$x = x' - y' \cot(\theta)$$
$$y = \frac{y'}{\sin(\theta)}$$

From 3D to 2D in matrix form

$$\mathbf{x} = \begin{bmatrix} m_x & 0 & x_0 \\ 0 & m_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\cot\theta & 0 \\ 0 & \frac{1}{\sin\theta} & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_c$$

$$= \begin{bmatrix} m_x & 0 & x_0 \\ 0 & m_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\cot\theta & 0 \\ 0 & \frac{1}{\sin\theta} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{R} \left[\mathbf{I} \mid -\mathbf{C} \right] \mathbf{X}$$

$$= \begin{bmatrix} m_x f & -m_x f \cot\theta & x_0 \\ 0 & \frac{m_y f}{\sin\theta} & y_0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{R} \left[\mathbf{I} \mid -\mathbf{C} \right] \mathbf{X}$$

$$= \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{R}[\mathbf{I} \mid -\mathbf{C}]\mathbf{X} = \mathbf{K}\mathbf{R}[\mathbf{I} \mid -\mathbf{C}]\mathbf{X} = \mathbf{P}\mathbf{X} \qquad (6)$$

Internal parameters

- m_x and m_y are the number of pixels per unit length in the skewed x-axis direction and skewed y-axis direction in image plane respectively
- f the focal length of camera
- x_0 and y_0 are the principal point coordinates in the skewed image frame (pixels)
- s the skewness factor which is 0 when the pixel is rectangle
- θ the skewness angle between two sides of image CCD plane

Calibration matrix K

K is called the internal calibration matrix. It is an intrinsic camera property:

$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} m_x f & -m_x f \cot\theta & x_0 \\ 0 & \frac{m_y f}{\sin\theta} & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

Observe that the entries in **K** are not all positive. θ is generally in the range $[0, \pi]$. The entry $-m_x f \cot\theta$ will be positive if $\theta > 90°$, negative if $\theta < 90°$ and 0 when $\theta = 90°$. The entry $\frac{m_y f}{\sin(\theta)}$ will always be positive. So the determinant of **K** will be always be positive.

Projection matrix

The central projection from 3D to 2D can be represented by a 3×4 matrix: $\mathbf{P} = \mathbf{K}\mathbf{R}\,[\mathbf{I} \mid -\mathbf{C}]$, which is called camera projection matrix. This matrix contains all the parameters of camera: calibration matrix as internal parameters; camera orientation and camera center as external parameters.
The projection matrix has always rank 3.

Projective geometry

- Projective geometry deals with the geometric properties that are invariant under projective transformations
- It is easier to understand projective geometry in 2D, which is in fact the geometry of projective transformations of the plane
- A 2D projective transformation arises when a plane is imaged by a pinhole camera
- Under perspective imaging certain geometric properties are preserved while others are not

Homogeneous coordinates

Homogeneous coordinates is very useful in multi-view geometry, which can easily represent many fundamental relationships in vector or matrix form.

A line in the plane can be represented by an equation $ax + by + c = 0$ with $(x, y)^T$ a point on the line.
In vector form: $\mathbf{x}^T \mathbf{l} = 0$ with $\mathbf{x} = (x, y, 1)^T$ and $\mathbf{l} = (a, b, c)^T$.

Any vectors $m(x, y, 1)^T$ and $n(a, b, c)^T$ also satisfy the line equation for any $m \neq 0$ and $n \neq 0$.
So two vectors related by an overall non-zero scaling are considered as being equivalent.

An equivalence class of vectors under this equivalence relationship

Homogeneous coordinates

For a point in the plane, its homogeneous coordinates is of the form $\mathbf{x} = (x_1, x_2, x_3)^T$, representing the point inhomogeneous coordinates $(x_1/x_3, x_2/x_3)^T$ $(x_3 \neq 0)$ in $\mathcal{R}^2$.

Even if the homogeneous coordinates of points and lines in the plane is a 3D vector, its degrees of freedom (DOF) are always 2.

Other properties

Given two lines $\mathbf{l} = (a, b, c)^T$ and $\mathbf{l}' = (a', b', c')^T$, the homogeneous coordinates of the intersection point is $\mathbf{x} = \mathbf{l} \times \mathbf{l}'$ with $\times$ the cross product:

$$\mathbf{l} \times \mathbf{l}' = \begin{pmatrix} bc' - b'c \\ ca' - c'a \\ ab' - a'b \end{pmatrix}$$

The line passing through two points $\mathbf{x}$ and $\mathbf{x}'$ has the form $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$.

Points at infinity

Given two parallel lines $\mathbf{l} = (a, b, c)^T$ and $\mathbf{l}' = (a, b, c')^T$, the intersection point of the two lines is $\mathbf{l} \times \mathbf{l}' = (b, -a, 0)^T$, corresponding to inhomogeneous coordinates $(b/0, -a/0)^T$.

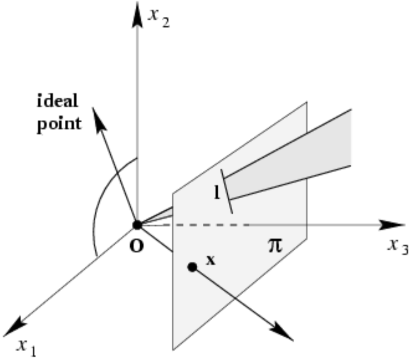
Any point with homogeneous coordinates $(x, y, 0)^T$ does not correspond to any finite point in $\mathcal{R}^2$.

Parallel lines meet at infinity.

2D projective space

- Homogeneous vectors $\mathbf{x} = (x_1, x_2, x_3)^T$ such that $x_3 \neq 0$ correspond to finite points in $\mathcal{R}^2$.
- By augmenting $\mathcal{R}^2$ with points having last coordinates $x_3 = 0$, the resulting space is the set of all homogeneous 3-vectors, namely the 2D projective space $\mathcal{P}^2$.
- The points with last coordinates $x_3 = 0$ are called ideal points or points at infinity. Each ideal point represents a direction determined by the ratio $x_1 : x_2$ $(x_2 \neq 0)$ or $x_2 : x_1$ $(x_1 \neq 0)$.
- All the ideal points lie on a line at infinity, denoted by $\mathbf{l}_\infty = (0, 0, 1)^T$. It can be verified that $(x_1, x_2, 0)(0, 0, 1)^T = 0$. Each line $\mathbf{l}$ intersects $\mathbf{l}_\infty$ at an ideal point, which corresponds to the direction of $\mathbf{l}$.

Projective plane



$$\mathcal{P}^2 = \mathcal{R}^3 - (0,0,0)^T$$

Transformations

The projective transformation (or homography), which is a non-singular 3×3 matrix, usually denoted by $\mathbf{H}$.

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (8)$$

$$\begin{aligned} \frac{x'_1}{x'_3} &= \frac{h_{11}x_1 + h_{12}x_2 + h_{13}x_3}{h_{31}x_1 + h_{32}x_2 + h_{33}x_3} \\ \frac{x'_2}{x'_3} &= \frac{h_{21}x_1 + h_{22}x_2 + h_{23}x_3}{h_{31}x_1 + h_{32}x_2 + h_{33}x_3} \end{aligned} \quad (9)$$

$\mathbf{H}$ is also a homogeneous geometric entity and it has 8 degrees of freedom.

Collinearity

A point $\mathbf{x}$ is transformed to point $\mathbf{Hx}$ under homography H, while a line $\mathbf{l}$ is transformed to a line $\mathbf{H}^{-\mathbf{T}}\mathbf{l}$.

Collinearity: if $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are on a the line $\mathbf{l}$, then $\mathbf{Hx}_1$, $\mathbf{Hx}_2$ and $\mathbf{Hx}_3$ are also on a the line $\mathbf{H}^{-T}\mathbf{l}$. More details about projective transformation can be found in

R.I. Hartley and A. Zisserman.
Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

Camera rotation

A 2D projective transformation can be induced by a pure camera rotation without changing its optic center.
Given a 3D point $\mathbf{X}$, its projected image by rotating the camera is:

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{K}_1\mathbf{R}_1[\mathbf{I} \mid -\mathbf{C}]\mathbf{X} \\ \mathbf{x}_2 &= \mathbf{K}_2\mathbf{R}_2[\mathbf{I} \mid -\mathbf{C}]\mathbf{X} \end{aligned} \quad (10)$$

$\mathbf{x}_1$ and $\mathbf{x}_2$ are related by a homography:

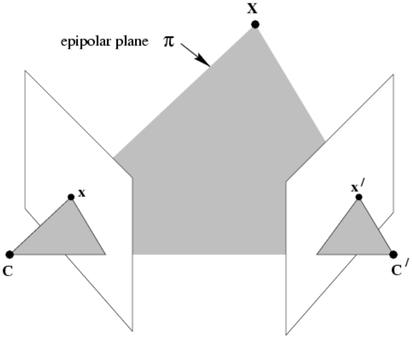
$$\mathbf{x}_2 = \mathbf{K}_2\mathbf{R}_2\mathbf{R}_1^{-1}\mathbf{K}_1^{-1}\mathbf{x}_1 = \mathbf{H}\mathbf{x}_1 \quad (11)$$

Homography

A homography can be induced from:

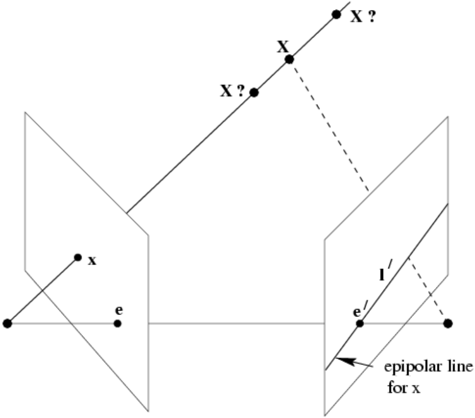
- A camera rotation without changing its optic center
- The 3D scene is a plane
- The scene is very far away from the camera.

Epipolar constraint



Terminology: epipolar plane, epipolar line, epipole

Epipolar constraint



$$\mathbf{x}'^T \mathbf{F}\mathbf{x} = 0$$

$$\mathbf{l}' = \mathbf{F}\mathbf{x}$$

Some properties of $\mathbf{F}$

- F is a 3×3 rank-2 homogeneous matrix with 7 freedom degrees
- $\mathbf{x}'^T \mathbf{F}\mathbf{x} = 0$ for a pair of corresponding image points $\mathbf{x}$ and $\mathbf{x}'$
  - Given a point $\mathbf{x}$ in the left image, the corresponding epipolar line in the right image is $\mathbf{l}' = \mathbf{F}\mathbf{x}$
  - Given a point $\mathbf{x}'$ in the right image, the corresponding epipolar line in the left image is $\mathbf{l} = \mathbf{F}^T\mathbf{x}'$
- $\mathbf{F}\mathbf{e} = \mathbf{0}$ and $\mathbf{F}^T\mathbf{e}' = 0$

Some properties of $\mathbf{F}$

- Computation of $\mathbf{F}$ from camera projection matrix $\mathbf{P}$ and $\mathbf{P}'$
  $\mathbf{F} = [\mathbf{e}']_\times \mathbf{P}'\mathbf{P}^+$ with $\mathbf{P}^+$ the pseudo-inverse of $\mathbf{P}$ and $\mathbf{e}' = \mathbf{P}'\mathbf{C}$
- Correspondence between epipolar lines
  $\mathbf{l}' = \mathbf{F}[\mathbf{e}]_\times \mathbf{l}$ and $\mathbf{l} = \mathbf{F}^T[\mathbf{e}']_\times \mathbf{l}'$

$$\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_\times \mathbf{b}, \text{ with } \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

$$\mathbf{a}_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

The rectified configuration

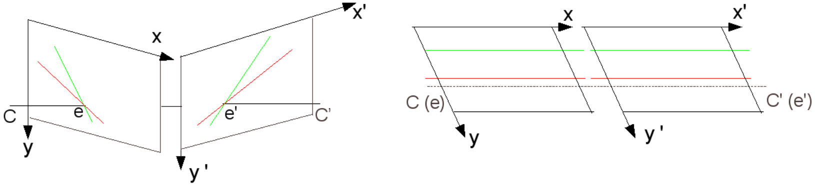


Figure: Left: General case. Right: Rectified case.

In the rectified case, the translation of the camera is parallel to its x-axis and there is no rotation.

Special form of $\mathbf{F}$

- General formula for $\mathbf{F}$:

$$\mathbf{F} = \mathbf{K}'^{-T}[\mathbf{T}]_\times \mathbf{R}\mathbf{K}^{-1}.$$

- Here, $\mathbf{R} = \mathbf{I}$ and $\mathbf{T} = -\lambda\mathbf{i}$:

$$\mathbf{F} = -\lambda \mathbf{K}^{-T}[\mathbf{i}]_\times \mathbf{K}^{-1}$$

- As $\mathbf{K}$ is upper-triangular, we get:

$$\mathbf{F} = [\mathbf{i}]_\times = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

- Epipolar constraint: $y' = y$.

Decomposition of F

- Suppose we can decompose **F** as:

$$\mathbf{F} = \mathbf{H}'^T [\mathbf{i}]_\times \mathbf{H}$$

- The images $u_g \circ \mathbf{H}'^{-1}$ and $u_d \circ \mathbf{H}^{-1}$ are rectified since:

$$\mathbf{y}'^T [\mathbf{i}]_\times \mathbf{y} = (\mathbf{H}'\mathbf{x})^T [\mathbf{i}]_\times (\mathbf{H}\mathbf{x}) = \mathbf{x}'^T \mathbf{F}\mathbf{x}.$$

- Invariance through rotation around the baseline:

$$\mathbf{R}_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c & -s \\ 0 & s & c \end{pmatrix} \Rightarrow \mathbf{R}_x^T [\mathbf{i}]_\times \mathbf{R}_x = [\mathbf{i}]_\times$$

- 3×3 equations, $8 + 8$ degrees of freedom: multiple solutions

Simulating rotations

- Assumption: $\mathbf{K} = \mathbf{K}'$ (same camera settings in both images)

$$\mathbf{K} = \begin{pmatrix} \alpha & 0 & w/2 \\ 0 & \alpha & h/2 \\ 0 & 0 & 1 \end{pmatrix}$$

- We look for rotation matrices $\mathbf{R}_l$ and $\mathbf{R}_r$ such that

$$\mathbf{F} = (\mathbf{K}\mathbf{R}_l\mathbf{K}^{-1})^T [\mathbf{i}]_\times (\mathbf{K}\mathbf{R}_r\mathbf{K}^{-1})$$

- This simplifies as:

$$\mathbf{F} = \mathbf{K}^{-T}\mathbf{R}_l^T [\mathbf{i}]_\times \mathbf{R}_r\mathbf{K}^{-1}$$

- Due to invariance through x-axis rotation, we can assume no x-rotation in $\mathbf{R}_l$.

The unknowns

- 5 angles:

$$\mathbf{R}_l = \mathbf{R}_z(\theta_{lz})\mathbf{R}_y(\theta_{ly}) \quad \mathbf{R}_r = \mathbf{R}_z(\theta_{rz})\mathbf{R}_y(\theta_{ry})\mathbf{R}_x(\theta_{rx})$$

- 1 scalar: α, the focal length.
- Actually, very different ranges: angles $\theta \in [-\pi, \pi]$ and $\alpha \in [(w+h)/3, (w+h) \times 3]$
- We take instead as unknown $\beta = \log_3(\alpha/(w+h)) \in [-1, 1]$

Measuring the errors

- We do not know how to decompose **F** as above
- Instead, we want to minimize the distance of each point to its epipolar line:

$$\sum_i (d^2(\mathbf{H}_l \mathbf{x}_{li}, [\mathbf{i}]_\times \mathbf{H}_r \mathbf{x}_{ri}) + d^2(\mathbf{H}_r \mathbf{x}_{ri}, [\mathbf{i}]_\times \mathbf{H}_l \mathbf{x}_{li}))$$

with d^2 the square point-line distance

Algebraic expression of error

- Instead, a simpler algebraic error is considered:

$$E_i^2 = \frac{(\mathbf{x}_{li}^T \mathbf{F} \mathbf{x}_{ri})^2}{\|\overline{\mathbf{F}\mathbf{x}_{ri}}\|^2 + \|\overline{\mathbf{F}^\mathbf{T}\mathbf{x}_{li}}\|^2}$$

with $\overline{\begin{pmatrix} a & b & c \end{pmatrix}^T} = \begin{pmatrix} a & b \end{pmatrix}^T$

- We minimize the sum of these terms with our expression of **F** depending on the 6 unknowns.

Derivatives with respect to parameters

- Let us write

$$E_i = \frac{\mathbf{x}_{li}^T \mathbf{F} \mathbf{x}_{ri}}{(\|\overline{\mathbf{F}\mathbf{x}_{ri}}\|^2 + \|\overline{\mathbf{F}^\mathbf{T}\mathbf{x}_{li}}\|^2)^{1/2}} = \frac{N}{D}$$

- Then given a parameter p,

$$\frac{1}{2}\frac{\partial E_i}{\partial p} = \frac{\mathbf{x}_{il}^T \mathbf{F}' \mathbf{x}_{ir}}{D} - N\frac{\overline{\mathbf{F}^T \mathbf{x}_{il}}^T \overline{\mathbf{F}'^T \mathbf{x}_{il}} + \overline{\mathbf{F}\mathbf{x}_{ir}}^T \overline{\mathbf{F}'\mathbf{x}_{ir}}}{D^3}$$

with $\mathbf{F}' = \frac{\partial \mathbf{F}}{\partial p}$

- We have to compute $\mathbf{F}'$ for each parameter.

Partial derivatives of F

- For a rotation:

$$\mathbf{R}_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \Rightarrow \mathbf{R}'_x(\theta) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\sin\theta & -\cos\theta \\ 0 & \cos\theta & -\sin\theta \end{pmatrix}$$

- For $\mathbf{K}$, we have

$$\mathbf{K}^{-1}(\alpha) = \begin{pmatrix} 1/\alpha & 0 & -w/(2\alpha) \\ 0 & 1/\alpha & -h/(2\alpha) \\ 0 & 0 & 1 \end{pmatrix}$$

so its derivative with respect to β:

$$\frac{\partial\mathbf{K}^{-1}}{\partial\beta} = \frac{\partial\mathbf{K}^{-1}}{\partial\alpha}\frac{\partial\alpha}{\partial\beta} = -\log 3 \begin{pmatrix} 1/\alpha & 0 & -w/(2\alpha) \\ 0 & 1/\alpha & -h/(2\alpha) \\ 0 & 0 & 0 \end{pmatrix}$$

Levenberg-Marquardt minimization

- We have $\mathbf{E} : \mathbb{R}^6 \to \mathbb{R}^n$ (n correspondences)
- Objective: find $\mathbf{x}$ that minimizes $\|\mathbf{E}(\mathbf{x})\|^2$
- If we write $\mathbf{E}(\mathbf{x}_0 + \Delta) = \mathbf{E}(\mathbf{x}_0) + \mathbf{J}\Delta$, minimize over Δ:

$$\|\mathbf{E}(\mathbf{x}_0) + \mathbf{J}\Delta\|^2 = \|\mathbf{E}(\mathbf{x}_0)\|^2 + 2(\mathbf{J}^T\mathbf{E}(\mathbf{x}_0))^T\Delta + \|\mathbf{J}\Delta\|^2$$

- Solution must satisfy the linear system: $(\mathbf{J}^T\mathbf{J})\Delta = -\mathbf{J}^T\mathbf{E}(\mathbf{x}_0)$.
- Augmented equation: $(\mathbf{J}^T\mathbf{J} + \lambda \text{diag}(\mathbf{J}^T\mathbf{J}))\Delta = -\mathbf{J}^T\mathbf{E}(\mathbf{x}_0)$
- If $\|\mathbf{E}(\mathbf{x}_0 + \Delta)\|^2 < \|\mathbf{E}(\mathbf{x}_0)\|^2$: iterate with $\mathbf{x}_0 \mathrel{+}= \Delta$, $\lambda \mathrel{/}= 10$
- If $\|\mathbf{E}(\mathbf{x}_0 + \Delta)\|^2 \geq \|\mathbf{E}(\mathbf{x}_0)\|^2$: iterate with same $\mathbf{x}_0$, $\lambda \mathrel{*}= 10$

Null columns of the Jacobian

- In equation $(\mathbf{J}^T\mathbf{J})\Delta = -\mathbf{J}^T\mathbf{E}(\mathbf{x}_0)$ we must have $\mathbf{J}$ of rank 6 so that $\mathbf{J}^T\mathbf{J}$ be invertible
- In particular, if some column of $\mathbf{J}$ is $\mathbf{0}$, we get a scalar equation $\mathbf{0}^T\Delta = 0$
- Solution: remove such equations from the system before solving.
- This happens for $\frac{\partial \mathbf{E}}{\partial \beta}$ at initial position $\mathbf{R}_l = \mathbf{R}_r = \mathbf{I}$ (column 6 of $\mathbf{J}$)

Summary of the rectification pipeline

1. Find correspondences between image pairs (SIFT)
2. Eliminate false correspondences by rigidity constraint (RANSAC searching for epipolar matrix)
3. Levenberg-Marquardt minimization of the error function
4. Apply homographies to images (pull values from initial images rather than push pixels to final image)
5. Then what? search for corresponding points reduced to horizontal direction

Ruins

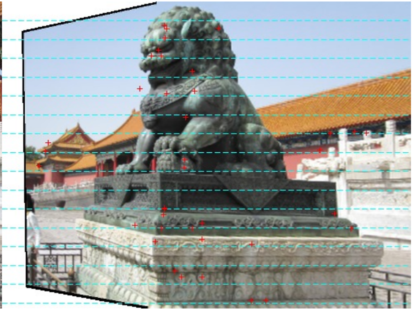
$||E_0|| = 3.21$ pixels.

$||E_6|| = 0.12$ pixels.

Ruins

$||E_0|| = 3.21$ pixels. $||E_6|| = 0.12$ pixels.

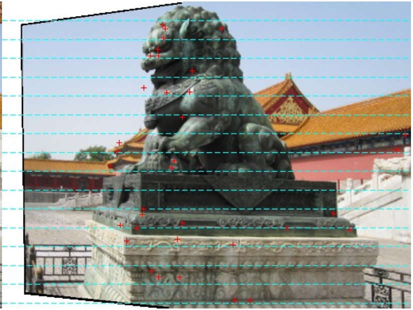
Beijing lion



$\|E_0\| = 4.32$ pixels.

$\|E_7\| = 0.36$ pixels.

Beijing lion



$||E_0|| = 4.32$ pixels.

$||E_7|| = 0.36$ pixels.

Cake

$||E_0|| = 17.9$ pixels.

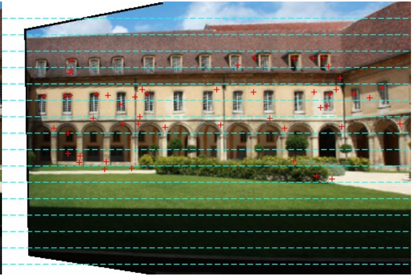
$||E_{13}|| = 0.65$ pixels.

Cake

$\|E_0\| = 17.9$ pixels.

$\|E_{13}\| = 0.65$ pixels.

Cluny



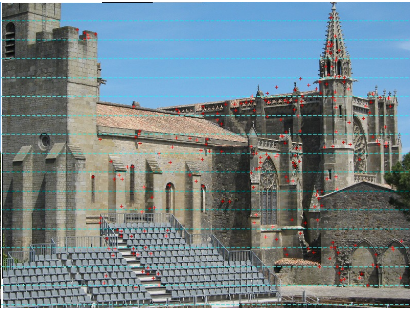
$\|E_0\| = 4.87$ pixels.

$\|E_{14}\| = 0.26$ pixels.

Cluny

$\|E_0\| = 4.87$ pixels. $\|E_{14}\| = 0.26$ pixels.

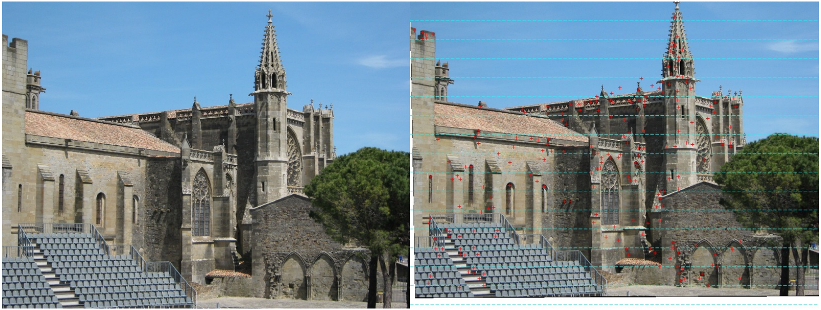
Carcassonne



$||E_0|| = 15.6$ pixels.

$||E_4|| = 0.24$ pixels.

Carcassonne



$||E_0|| = 15.6$ pixels.

$||E_4|| = 0.24$ pixels.

Books

$||E_0|| = 3.22$ pixels. $||E_{14}|| = 0.27$ pixels.

Books

$||E_0|| = 3.22$ pixels.

$||E_{14}|| = 0.27$ pixels.

Project: Hartley's method (1999)

R.I. Hartley.
Theory and practice of projective rectification. *International Journal of Computer Vision*, 35(2):115–127, 1999.

- Compute **F** from point correspondences
- Rotate image and send epipole to infinity in x direction
- Apply affine transform $x' = ax + by + c$ so as to minimize disparities

Project: Gluckman-Nayar (2001)

J. Gluckman and S.K. Nayar.
Rectifying transformations that minimize resampling effects.
IEEE Conf. Computer Vision and Pattern, 1:111, 2001.

- Local area change causes loss or creation of pixels
- Area change measured by $\det(\mathbf{J})$, $\mathbf{J}$ being the Jacobian matrix of $\mathbf{H}$.
- Minimize *w.r.t.* 2 variables the distortion $E(\mathbf{H}) + E(\mathbf{H}')$ with

$$E(\mathbf{H}) = \iint \left(\det \left(\frac{\partial \mathbf{H}(x,y)}{\partial(x,y)} \right) - 1 \right)^2 dx\, dy$$

- Rational polynomial of degree 16 for one variable, quadratic for the other

Loop-Zhang (1999)

C. Loop and Z. Zhang.
Computing rectifying homographies for stereo vision.
Computer Vision and Pattern Recognition, 1:125–131, 1999.

- 3 parts: projective, similarity, shear, each minimizing the distortion
- Projective: find a transform that sends $\mathbf{e}$ to infinity and keeps a point $\mathbf{z} \in \mathbf{l}_\infty$ fixed. 7-order polynomial root extraction to find $\mathbf{z}$.
- Similarity: send epipole to $\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^T$
- Shear: already rectified case, but tries to keep orthogonality of middle lines.